Data & Information

a) [1 point] What color is the flower in `flower-bsq-u8be-3x1512x2268.raw`? (dimensions are expressed as $Z \times Y \times X$, u8 indicates "unsigned, 8-bit integer samples", and "be" denotes "Big endian").

The average RGB value: R:92.61 G:45.38 B:72.80,
The flower is purple (red dominates, with blue distinctly present, and some green present),
However, I caculate the average color intensity across the entire image, it is evident that the surrounding green environment influnces the result of "G", so flower itself will have lower G value.

b) [2 points] What's the information rate of that flower, and of each of its RGB components? Round to 2 decimals.

What	Information rate (bits per sample)
The whole flower	7.09 bits / sample
Its red component	7.28 bits / sample
Its green component	5.89 bits / sample
Its blue component	7.26 bits / sample

c) [1 point] One of these is pretty different from the others. How would you explain it?

This is because the dominant colors in the image are primarily red and blue, with minimal green components. Consequently, the green channel carries less information, resulting in lower entropy.

d) [2 points] We want to store the original flower in PPM format. This format consists of a header (3 lines of text) immediately followed by binary data. The file `flower_ppm_header.txt` contains the necessary header. After the header, pixels are stored in BIP order (band interleaved by pixel) with the same bitdepth as the original (8 bps). Store the result in a file with `.ppm` extension and open it with your favorite image viewer. You should see the same flower.

Let H_{raw} be the information rate of the original `.raw` file, and H_{ppm} that of the new PPM file, both in bits per sample. Express the difference $H_{\text{raw}} - H_{\text{ppm}}$ in scientific notation with 2 decimals of precision. Explain why this value is or isn't exactly zero.

$H_{raw} - H_{ppm(data)} = 0.00$ bits/sample. → The pixel data in RAW files and PPM files are same, differing only in storage order (BSQ, BIP) so information content remains unchanged.

$H_{raw} - H_{ppm(full)} \approx -6.80 \times 10^{-7}$ bits/sample → PPM files append a small text header, slightly changing the overall byte distribution and thereby introducing a minute deviation in entropy.

(turn the page to continue...)

e) **[2 points]** Finally, we want to save the flower in two files. Both will have raw format and use 2 bytes per sample, but only one of them can be big endian. In addition, the first and last samples of these two files must be set to `0x0000` and `0xffff`, respectively.

After creating the two files, open the original and the multibyte flowers with `imagej/fiji`. Explain why the histograms (visible by pressing `Ctrl+H` should be identical).

The histograms of Big-endian and small-endian Raw files are same.

Reason: Endianness merely alters the byte order without changing the values themselves; the mapping from 8-bit to 16-bit values is a linear scaling, preserving the probality distribution, so the infurmain content and entropy remain consistent.

---

f) **[2 points]** Let H_{raw} be the information rate of the original `.raw` file, as before. Let H_{BE} and H_{LE} be the multibyte, big-endian and little-endian samples. Express the differences $H_{\text{raw}} - H_{\text{BE}}$ and $H_{\text{raw}} - H_{\text{LE}}$ in scientific notation with 2 decimals of precision, and explain why these values are or aren't exactly zero.

$H_{raw} - H_{BE} \approx -1.48 \times 10^{-6}$ bits/sample

$H_{raw} - H_{LE} \approx -1.48 \times 10^{-6}$ bits/sample.

Theoretically, it should be zero, as the mapping does not alter the information content; however, the forced modification of the first and last samples slightly alters the pixel distribution, resulting in a minimal deviation in entropy.